SUPPLEMENTARY TABLES AND FIGURES

Supplementary Table S1: Summary of CNV characteristics used in our analysis.

Antimalarial	Parent clone	Clones	Putative gene amplified (chromosome)	Amp. sizes	Data source	Accession reference
DSM1	Dd2	C	Dihydroorotate dehydrogenase (6)	~70kb	[1]	SRX326516
		D		~95kb		SRX326519
		E		~34kb		N/A
		F		~39kb		N/A
		Parent		N/A		SRX326518
Halofuginone	Dd2	HFGRII	Prolyl–tRNA synthetase (12)	~30kb	[2]	SRX158283
		HFGRIII				SRX200273
		Parent		N/A		SRX738616
MMV029272	3d7	R2B2	ABC transporter I family member, putative (1)	~62kb	[3]	SRX2479359
		R2C9				SRX2479247
		R3E7				SRX2479252
		R3F10				SRX2479375
		Parent		N/A		SRX2479354
MMV019662	3d7	1C4	Lipid/sterol:H+ symporter (1)	~95kb		SRX2479223
		2B6				SRX2479224
		2F6				SRX2479226
		3-G6				SRX2479265
		F7				SRX2479256
		2D6				SRX2479372
		3B6		~99kb		SRX2479268
		1F4		~52kb		SRX2479340
		1F9				SRX2479347
		33XC3		~35kb		SRX2479331
		3C3		~51kb		SRX2479355
		3F10				SRX2479219
		2G6		~41kb		SRX2479399
		2G9				SRX2479357
		Parent		N/A		SRX2479243
MMV028038	3d7	2E3	Lipid/sterol:H+ symporter (1)	~51kb		SRX2479393
		2F10				SRX2479204
		3E9				SRX2479235
		3F5				SRX2479392
		1E10		~41kb		SRX2479244
		1E3				SRX2479242
		Parent		N/A		SRX2479243
MMV08149	Dd2	1B2	Unknown (10, 12)	~18kb, ~30kb		SRX1561330
		Parent		N/A		SRX5161067
Cladosporin	Dd2	CladoA	Lysyl tRNA Synthetase, (13)	~58kb		SRX2479289
		CladoB		~50kb		SRX2479338
		CladoC		~35kb		SRX2479378
		Parent		N/A		SRX2479309
Primaquine	Dd2	PQA11	Patatin-like phospholipase, putative (10)	~18kb		SRX2479288
		Parent		N/A		SRX2479263

N/A = whole genome sequencing not available (For DSM1 clones, CNVs were determined by PCR across breakpoints and microarrays).

Supplementary Table S2: Alignment statistics and mapping quality.

Clone	# of mapped reads	% mapped of total reads	Mean coverage (reads/bp ± std. dev)	Mean Mapping Quality*	Median Insert Size*	Mean Coverage 2kb around breakpoint regions (reads/bp ± std. dev)	Mean Coverage 100bp around breakpoint regions (reads/bp ± std. dev)
DSM1-C	23,606,598	98.73	96.2 ± 68	57.0	308	465.8 ± 435.6	315.4 ± 249.9
DSM1-D	58,986,651	97.95	210.2 ± 121.7	54.82	261	872.3 ± 737.1	125,592.0 ± 256,594.6
HFGRII	35,595,585	98.02	143.9 ± 63.4	56.4	144	158.2 ± 37.9	97.1 ± 25.0
HFGRIII	35,477,690	98.18	142.4 ± 91.6	54.46	150	255.4 ± 105.2	180.0 ± 65.9
CladoA	21,978,885	100	54.7 ± 34.8	55.4	321	118.5 ± 94.2	58.8 ± 30.9
CladoB	31,695,884	100	79.5 ± 54.2	55.6	349	152.3 ± 120.4	127.2 ± 68.9
CladoC	39,472,609	100	99.3 ± 64.9	55.8	294	237.2 ± 209.0	117.0 ± 89.0
PQA11	11,056,363	100	47.8 ± 76.5	57.0	267	60.7 ± 26.5	45.9 ± 22.3
F7	26,916,655	100	111.5 ± 141.4	57.8	242	127..0 ± 70.9	73.2 ± 33.4
3B6	15,482,302	100	63.5 ± 331.6	57.8	238	60.1 ± 49.7	30.3 ± 12.0
1F4	20,833,721	100	85.0 ± 95.9	57.8	227	118.1 ± 65.8	55.9 ± 33.3
2G9	15,622,275	100	61.6 ± 77.2	57.8	182	81.3 ± 52.2	48.4 ± 29.5
1E3	28,662,532	100	106.7 ± 127.2	57.7	122	117.0 ± 85.9	63.8 ± 39.9
33XC3	20,214,893	100	80.1 ± 74.7	57.6	160	67.7 ± 31.2	60.8 ± 16.8
3C3	24,656,913	100	97.1 ± 139.9	57.8	152	84.0 ± 111.1	38.59 ± 25.5
R2B2	21,492,697	100	89.2 ± 184.0	57.7	238	114.7 ± 72.3	23.8 ± 11.0
1B2ch10	24,513,373	90.23	85.9 ± 97.3	58.4	250	104.2 ± 61.8	62.5 ± 20.3
1B2ch12	24,513,373	90.23	85.9 ± 97.3	58.4	250	94.9 ± 78.7	39.7 ± 9.5
Clones with non-unique CNVs	**# of mapped reads**	**% mapped of total reads**	**Mean coverage (reads/bp ± std. dev)**	**Mean Mapping Quality***	**Median Insert Size**	**Mean Coverage 2kb around breakpoint regions (reads/bp ± std. dev)**	**Mean Coverage 100bp around breakpoint regions (reads/bp ± std. dev)**
R2C9	19,750,338	100	81.7 ± 179.8	57.7	255	127.7 ± 89.9	24.3 ± 13.5
R3E7	21,127,436	100	72.0 ± 75.1	57.8	213	107.3 ± 64.5	28.2 ± 13.7
R3F10	27,855,320	100	92.6 ± 97.3	57.6	160	141.0 ± 95.4	26.2 ± 12.8
1C4	13,804,219	100	56.9 ± 94.3	56.3	220	78.0 ± 67.3	73.6 ± 47.0
2B6	21,906,529	100	91.1 ± 127.2	57.8	253	99.2 ± 69.2	57.7 ± 30.9
2F6	25,244,172	100	104.6 ± 120.4	57.8	222	119.6 ± 68.9	84.7 ± 29.1
3G6	21,155,622	100	87.2 ± 303.6	57.8	223	90.3 ± 53.3	69.0 ± 26.3
2D6	20,427,288	100	84.9 ± 88.0	57.9	247	91.6 ± 62.5	51.7 ± 19.3
1F9	33,525,957	100	132.0 ± 178.8	57.8	159	149.6 ± 85.9	77.4 ± 43.9
3F10	21,739,212	100	80.5 ± 64.5	57.8	176	107.8 ± 115.7	45.0 ± 17.8
2E3	11,170,025	100	87.4 ± 101.3	57.7	128	21.5 ± 22	11.2 ± 2.8
2F10	24936046	100	86.7 ± 55.6	57.7	144	104.4 ± 96.3	52.2 ± 20.8
3E9	27,252,221	100	92.9 ± 70.6	57.6	132	98.0 ± 95.9	57.8 ± 29.7
3F5	29,612,259	100	114.6 ± 138.7	57.7	141	107.4 ± 97.8	53.87 ± 25.4
2G6	20,240,247	100	80.2 ± 104.4	57.8	193	114.0 ± 60.4	63.9 ± 42.1
1E10	24,613,192	100	91.2 ± 114.9	57.6	122	98.3 ± 71.2	56.8 ± 31.5

*Mean mapping quality was determined excluding 50kb from each end of chromosomes to avoid telomeric DNA, max value is 60. Median insert sizes are the median distance between mapped forward and reverse reads.

Supplementary Table S3: Variant statistics and confidence.

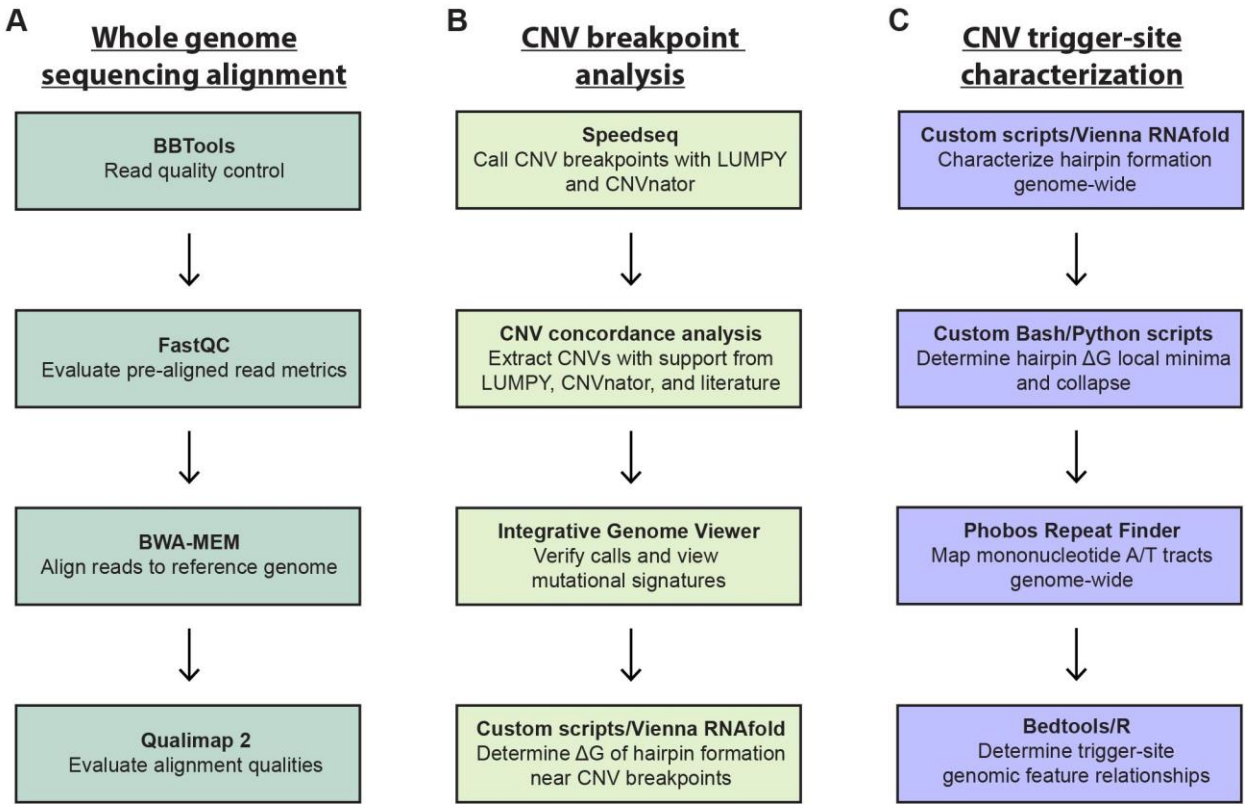
Clone	Orientation of amplification	LUMPY Sample Quality	LUMPY PE/SR Support	CNVnator Start	CNVnator End	CNVnator Copy #
DSM1-C	Tandem	18620.94	1025/0	79101	152500	7.2
DSM1-D	Tandem	8595.33	32/0	64501	158200	5.8
HFGRII	Inverted	174.29	3/0	N/A	N/A	N/A
HFGRIII	Tandem	902.33	44/0	575001	621900	2.0
CladoA	Tandem	2257.46	129/0	2000301	2058400	5.3
CladoB	Tandem	5542.93	330/0	2005701	2055100	5.0
CladoC	Tandem	7587.09	445/0	2000201	2022800	5.0
PQA11	Tandem	957.11	59/0	290001	308800	2.9
F7	Tandem	700.79	29/4	264301	359400	2.0
3B6	Tandem	338.98	39/3	264301	359300	2.2
1F4	Tandem	1574.1	11/0	321501	372900	2.3
2G9	Tandem	964.9	39/3	321501	360300	2.4
1E3	Tandem	1221.36	48/1	321601	362600	2.2
33XC3	Tandem	143.15	13/1	1733601	1768700	2.1
3C3	Tandem	307.73	7/1	1726001	1767900	2.4
R2B2	Tandem	179.06	12/0	782801	857600	2.1
1B2ch10	Tandem	231.77	22/0	285701	315700	2.4
1B2ch12	Tandem	528.74	33/0	1549901	1567200	2.3
Supporting Clones	**Orientation of amplification**	**LUMPY Quality Score**	**LUMPY PE/SR Support**	**CNVnator Start**	**CNVnator End**	**CNVnator Copy #**
R2C9	Tandem	459.26	24/0	783001	857600	3.0
R3E7	Tandem	241.39	15/0	782901	856300	2.1
R3F10	Tandem	208.27	11/0	783001	857600	2.0
1C4	Tandem	707.66	29/0	266201	359400	2.0
2B6	Tandem	709.47	30/1	264401	356400	2.1
2F6	Tandem	467.2	19/2	264301	359400	2.1
3G6	Tandem	437.43	17/1	266201	359400	2.1
2D6	Inverted	443.03	19/3	266201	359300	2.0
1F9	Tandem	1766.92	74/0	321601	372900	2.2
2G6	Tandem	1323.84	56/1	321601	364800	2.3
1E10	Tandem	969.48	38/1	321601	342600	2.2
3F10	Tandem	351.89	5/2	1718201[#]	1770000[#]	2.2
2E3	Tandem	474.66	5/1	1718201[#]	1768000[#]	2.0
2F10	Tandem	106.51	5/0	1718201	1768000	2.1
3E9	Tandem	419.7	7/1	1718201	1768100	2.0
3F5	Tandem	548.68	3/1	1718201	1768100	2.0

Amplification orientation was determined by comparing paired-end sequencing read-mate orientation and strand (Fig. S1). LUMPY sample qualities have no theoretical maximum but >100 are considered high quality calls. PE/SR= paired-end and split-read support respectively. CNVnator was unable to call read-depth analysis but visual inspection of bam file showed increase in coverage indicating presence of CNV. [#]Clones had contiguous duplication calls from CNVnator that were combined for the overall amplification.

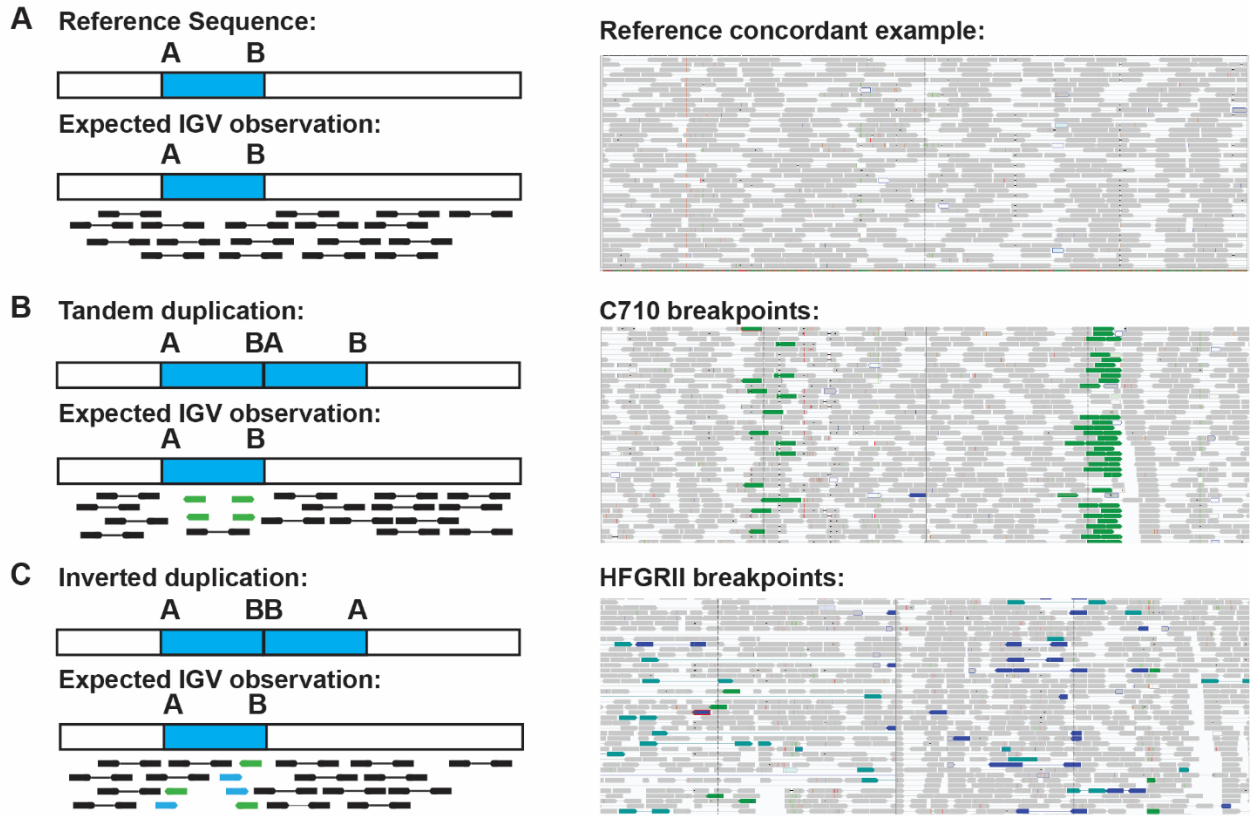
Supplementary Table S4: Comparison of A/T track breakpoint length pre- and post-CNV formation.

Shared Breakpoint	Pre-CNV A/T track length (bp)	Post-CNV A/T track length (bp)	% change	# of supporting split-reads	Mean phred score of split-read bases
DSM1F/C_3	37	31	-16	30	60
CladoA/C_5	40	ND	ND	ND	ND
F7/3B6_5	24	29	+21	2	60
1F4/1E3_5	33	29	-12	1	60
3B6/1E3_3	35	29	-18	3	60
Average	**34**	**30**	**-6**	**14**	**60**
Unique Breakpoint	**Pre-CNV A/T track length (bp)**	**Post-CNV A/T track length (bp)**	**% change**		
DSM1C_5	29	31	+7	30	60
DSM1D_5	38	20	-47	1	60
DSM1D_3	28	20	-29	1	60
DSM1E_5@	21	15	-29	ND	ND
DSM1E_5@	36	15	-58	ND	ND
DSM1-F@	32	25	-22	ND	ND
HFGRII_5	31	ND	ND	ND	ND
HFGRII_3	N/A^	N/A^	N/A^	N/A^	N/A^
HFGRIII_5	41	31	-24	2	60
HFGRIII_3	41	31	-24	2	60
CladoA_3	32	ND	ND	ND	ND
CladoB5	40	ND	ND	ND	ND
CladoB3	38	ND	ND	ND	ND
CladoC3	27	ND	ND	ND	ND
PQA11_5	37	26	-30	10	60
PQA11_3	26	26	0	10	60
1F4_3	25*	ND	ND	ND	ND
2G9_5	33	29	-12	3	60
2G9_3	35	29	-18	3	60
33XC3_5	N/A^	N/A^	N/A^	N/A^	N/A^
33XC3_3	N/A^	N/A^	N/A^	N/A^	N/A^
3C3_5	19	18	-5	3	60
3C3_3	34	18	-47	3	60
R2B2_5	24	ND	ND	ND	ND
R2B2_3	26	ND	ND	ND	ND
1B2ch10_5	35	ND	ND	ND	ND
1B2ch10_3	N/A^	ND	ND	ND	ND
1B2ch12_5	30	29	-3	3	60
1B2ch12_3	24	29	+21	3	60
Average	**32**	**25**	**-17**	**7**	**60**

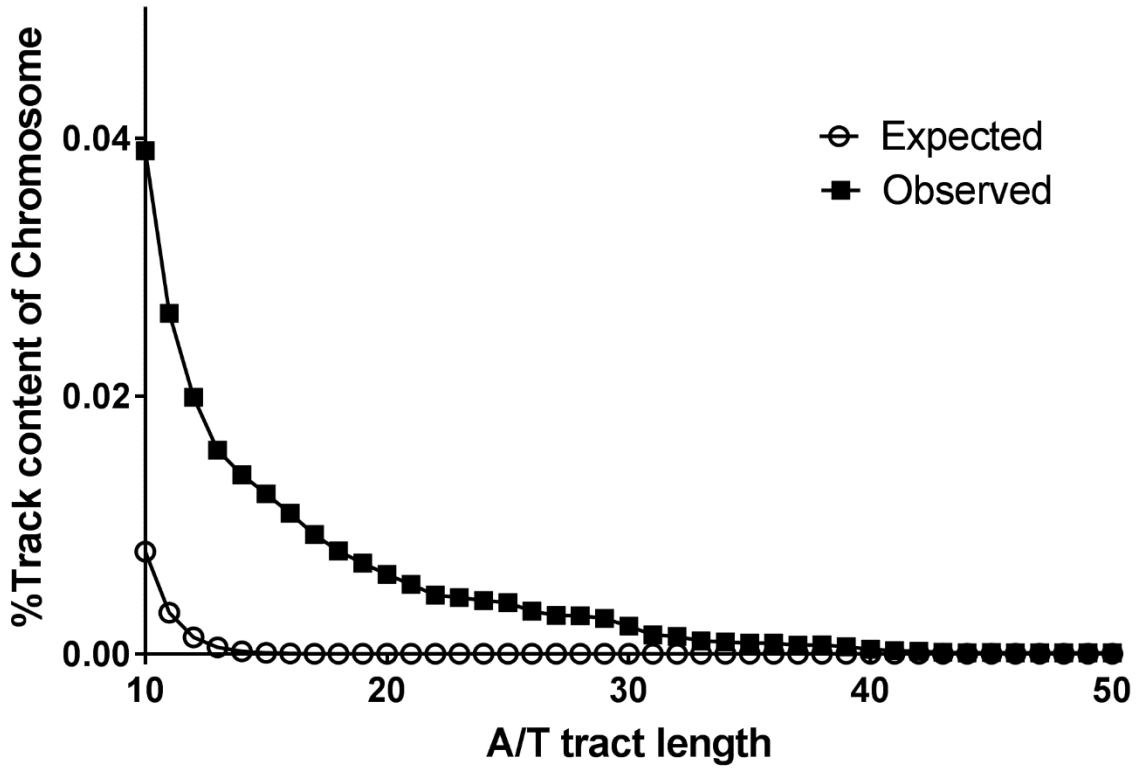
Post-CNV A/T track length was determined through split-reads from whole genome sequencing data. ND = not determined due to absence of split-reads mapped across breakpoints. N/A^ = AT dinucleotide repeats instead of A/T tracks, * = imperfect A/T track repeat



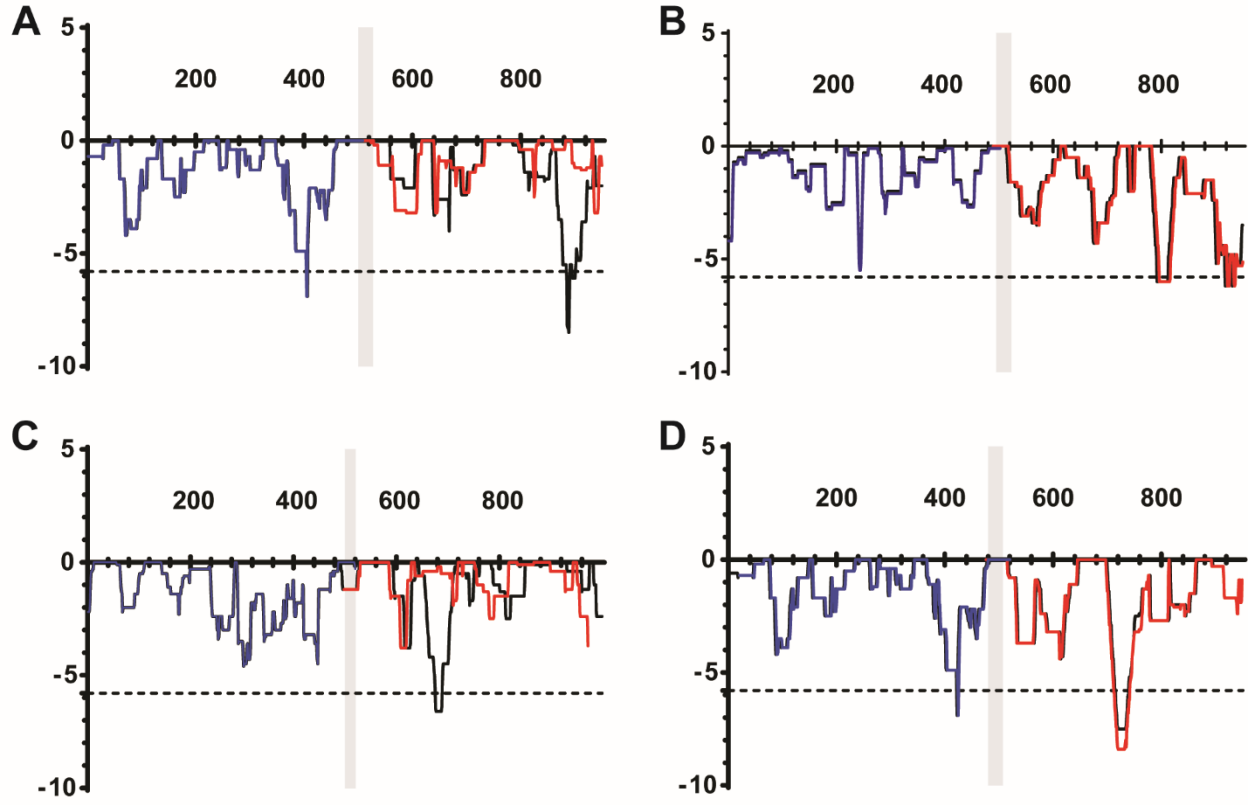
Supplementary Figure S1: Bioinformatic analysis of *Plasmodium* CNVs. **A.** Alignment of whole genome sequencing reads starts with BBTools to remove low quality bases or adapter sequences and verify correct pairing of reads. The resulting "clean" paired reads are evaluated by FastQC for overrepresented sequences, per base read qualities, and read length distributions. After passing read quality control, BWA-MEM is used to align "clean" paired reads to the *3d7 Plasmodium falciparum* reference genome. Qualimap 2 is then used to evaluate the alignments for mean/median read depth, paired read insert distributions, and mapping quality. **B.** After passing mapping quality control, Speedseq is used to call structural variants and CNVs with support from LUMPY, CNVnator, and positions from previous reports. The Integrative Genome Viewer is then used to manually verify CNV calls and evaluate mutational signatures such as read-pair orientation, CNV breakpoint sequences (i.e. A/T tract length), and proximal sequence changes that arise during CNV formation. Sequences windows around verified CNV breakpoints are extracted using a combination of custom Bash and Python scripts to create 50bp sliding windows with a 1bp shift and submitted to Vienna RNAfold for stable hairpin prediction. **C.** For genome-wide analysis, Vienna RNAfold is used to evaluate hairpin formation across all chromosomes (excluding subtelomeric/telomeric regions 50kb from the ends). Custom Bash/Python scripts are used to find local hairpin minima to find "stable hairpin forming regions". Phobos Repeat Finder is used on the same sequences to map mononucleotide A/T tracts. After mapping mononucleotide A/T tracts and stable hairpin forming regions, Bedtools and R are used to determine trigger-site feature relationships.



Supplementary Figure S2: Discordant read orientation of duplications. **A.** Reads aligning to the reference genome are colored based on read orientation and shown as pairs in IGV version 2.4.10. If reads match the reference sequence, they are expected to be gray and face towards each other as in the reference concordant example. **B.** If reads are found in a tandem duplication with respect to the reference sequence, they are colored green and face away from each other as in the C710 breakpoint example. These reads are shown with their pairs at their respective breakpoints and the insert sizes correspond to the size of the duplication. **C.** If reads are found in an inverted duplication with respect to the reference sequence, they are colored both blue and teal and are found facing each other and overlapping.



Supplementary Figure S3: Expected vs observed frequency of long A/T tracks. Frequency of (# tracks observed/chromosome length) for varying A/T tract lengths on all chromosomes. For equations used in calculation, see *Materials and Methods*.



Supplementary Figure S4: Post-CNV junctions indicate action of distinct repair pathways. Hairpin stability (ΔG) across 1kb of sequence at novel junctions created by the generation of CNVs (see **Fig. 4B**). Red and blue lines indicate *predicted* error-free repair utilizing pre-CNV sequence, black lines demark *observed* post-CNV sequence. Conserved junctions from D and F clones (panels **B** and **D**, respectively) indicate MMEJ action (see Fig. 4. Novel junctions created post-CNV rom C and E clones (panels **A** and **C**, respectively) indicate MMBIR action (also see **Fig. 4C** and **D**). Significant hairpins fall below the dotted black line (see methods for details on cut-off, -5.8 kCal/mol). The location of the A/T track at upstream and downstream breakpoints are indicated with vertical grey bars.

SUPPLEMENTARY REFERENCES

1. Guler, J.L., Freeman, D.L., Ahyong, V., Patrapuvich, R., White, J., Gujjar, R., Phillips, M.A., DeRisi, J. and Rathod, P.K. (2013) Asexual populations of the human malaria parasite, Plasmodium falciparum, use a two-step genomic strategy to acquire accurate, beneficial DNA amplifications. *PLoS Pathog*, **9**, e1003375.
2. Herman, J.D., Rice, D.P., Ribacke, U., Silterra, J., Deik, A.A., Moss, E.L., Broadbent, K.M., Neafsey, D.E., Desai, M.M., Clish, C.B. *et al.* (2014) A genomic and evolutionary approach reveals non-genetic drug resistance in malaria. *Genome Biology*, **15**, 511.
3. Cowell, A.N., Istvan, E.S., Lukens, A.K., Gomez-Lorenzo, M.G., Vanaerschot, M., Sakata-Kato, T., Flannery, E.L., Magistrado, P., Owen, E., Abraham, M. *et al.* (2018) Mapping the malaria parasite druggable genome by using in vitro evolution and chemogenomics. *Science*, **359**, 191-199.